THE BLOCK GRADE OF A BLOCK KRYLOV SPACE

MARTIN H. GUTKNECHT* AND THOMAS SCHMELZER†


Abstract. The aim of the paper is to compile and compare basic (theoretical) facts on Krylov subspaces and block Krylov subspaces. Many Krylov (sub)space methods for solving a linear system \( Ax = b \) have the property that in exact computer arithmetic the true solution is found after \( \nu \) iterations, where \( \nu \) is the dimension of the largest Krylov subspace generated by \( A \) from \( r_0 \), the residual of the initial approximation \( x_0 \). This dimension is called the grade of \( r_0 \) with respect to \( A \).

Though the structure of block Krylov subspaces is more complicated than that of ordinary Krylov subspaces, we introduce here a block grade for which an analogous statement holds when block Krylov space methods are applied to linear systems with multiple right-hand sides. The possibility of linear dependence among columns of the block Krylov matrix \( (r_0, A r_0, \ldots, A^{\nu-1} r_0) \), which in practical algorithms calls for the deflation of some columns, requires extra care. Relations between grade and block grade are also established, as well as relations to the corresponding notions of a minimal polynomial and its companion matrix.

Key words. sparse linear systems, multiple right-hand sides, several right-hand sides, block Krylov space method, block Krylov space solver, block size reduction, deflation, grade, minimal polynomial

AMS subject classifications. 65F10

1. Krylov spaces and their grade. Most currently used iterative methods for solving nonsingular linear systems of equations \( Ax = b \) are Krylov space solvers, which generate a sequence of approximate solutions \( x_n \), also called iterates, which are chosen from an affine space that grows with \( n \):

\[
x_n - x_0 \in K_n := K_n(A, r_0) := \text{span}(r_0, A r_0, \ldots, A^{n-1} r_0),
\]

where \( r_0 := b - A x_0 \) is the initial residual and \( K_n \) is the \( n \)th Krylov subspace generated by \( A \) from \( r_0 \). There are some Krylov space solvers, where \( x_n \) may not exist for some exceptional values of \( n \); but even in this case the results we cite here remain true as they are results about the subspaces, not individual methods.

We assume \( A \in \mathbb{C}^{N \times N} \), \( b \in \mathbb{C}^N \), \( x \in \mathbb{C}^N \) in this section, so \( K_n \subseteq \mathbb{C}^N \). Although Krylov space solvers are iterative in spirit, many of them actually produce in exact arithmetic the exact solution in at most \( N \) steps, and, as we see in a moment, this bound can then be replaced by a well determined integer that depends on \( A \) and \( b \), but, under a weak assumption, not on the method.

In practice this theoretical bound on the number of iterations may not take effect because the bound may be large and iterative methods are often strongly contaminated by roundoff. Nevertheless the basic facts on Krylov spaces deserve to be better known among the many users of Krylov space methods, and so do the basic facts on block Krylov spaces.

By definition, for given \( A \) and \( r_0 \), the subspaces \( K_n \) are clearly nested, and \( K_n \) can have at most dimension \( \min\{n, N\} \). However, one can say more. The following

*Seminar for Applied Mathematics, ETH-Zurich, CH-8092 Zurich, Switzerland; Email: mhg@math.ethz.ch, URL: http://www.sam.math.ethz.ch/~mhg.
†Oxford University Computing Laboratory, Oxford University, Wolfson Building, Parks Road, Oxford, OX1 3UQ, United Kingdom; Email: thoms@comlab.ox.ac.uk, URL: http://web.comlab.ox.ac.uk/oucl/people/thomas.schmelzer.html
lemmas and corollaries are well known and easy to prove, so we just outline the proofs. For generality and simplicity we replace $r_0$ by $y \neq o$ when appropriate.

**Lemma 1.** There is a positive integer $\nu := \nu(A, y)$ such that

$$\dim K_n(A, y) = \begin{cases} n & \text{if } n \leq \nu, \\ \nu & \text{if } n > \nu. \end{cases}$$

The inequalities $1 \leq \nu \leq N$ hold, and $\nu < N$ is possible if $N > 1$.

**Proof.** By definition, $\dim K_1 = 1$, and then for $n = 2, 3, \ldots$ we have $\dim K_n = n$ until we reach some $n = \nu$ where $A^n y$ is a linear combination of lower powers:

$$A^n y = -y \gamma_0 - Ay \gamma_1 - \cdots - A^{n-1} y \gamma_{n-1}. \quad (1.2)$$

Here, $\gamma_0 \neq 0$ but otherwise we could multiply by $A^{-1}$ and replace $\nu$ by $\nu - 1$. For $n > \nu$ we can then clearly also write all terms $A^n y$ as linear combinations of $y, Ay, \ldots, A^{n-1} y$.

To see that $\nu < N$ is possible we choose $y$ as an eigenvector. □

The relation (1.2) can be reformulated as

$$\psi(A)y = o, \quad \text{where} \quad \psi(t) := \psi_{A,y}(t) := t^{\nu} + \gamma_{\nu-1} t^{\nu-1} + \cdots + \gamma_1 t + \gamma_0. \quad (1.3)$$

Following Wilkinson [18] we make the following definition.¹

**Definition.** The positive integer $\nu := \nu(A, y)$ of Lemma 1 is called *grade of $y$ with respect to $A$* (or, *grade of $A$ with respect to $y$*). The polynomial $\psi_{A,y}$ of (1.3) is the *minimum polynomial of $y$ with respect to $A$*. ▲

There is no generally accepted preference for using the term “Krylov space” or the term “Krylov subspace”, but one could argue that $K_\nu(A, y)$ is the Krylov space while $K_n(A, y)$ with $n \leq \nu$ are Krylov subspaces.²

Lemma 1 provides us with a first characterization of $\nu(A, y)$:

**Corollary 2.** The grade $\nu(A, y)$ satisfies

$$\nu(A, y) = \min \{ n \mid \dim K_n(A, y) = \dim K_{n+1}(A, y) \} = \min \{ n \mid K_n(A, y) = K_{n+1}(A, y) \}. \quad (1.3)$$

There are several further characterizations of $K_\nu(A, y)$ and of the polynomial $\psi_{A,y}$. First, here is one for each.

**Lemma 3.** $K_\nu(A, y)$ is the smallest $A$–invariant subspace that contains $y$. Among the polynomials $\psi$ satisfying $\psi(A)y = o$, the polynomial $\psi = \psi_{A,y}$ has the smallest degree. In particular, $\nu$ is bounded by the degree of the minimal polynomial $\hat{\chi}_A$ of $A$.

---

¹Algebraists call $K_\nu(A, y)$ an $A$–cyclic (sub)space or the cyclic $C[x]$–submodule induced by $A$ and generated by $y$; see, e.g., page 356 of [8] or pages 146–147 of [14]. Based on that, Ilić and Turner [9] call $\nu$ the algebraic grade of $A$ with respect to $y$. Unlike numerical analysts, algebraists seem mainly interested in the cases where either $K_\nu(A, y)$ is an invariant subspace that belongs to a single Jordan block (i.e., to a single elementary divisor) or it is the whole space (i.e., here $C^N$). In the latter case the starting vector $y$ is called an $A$–cyclic vector or a cyclic vector for $A$. Algebraists call $\psi_{A,y}$ the $A$–annihilator of $y$, and the same name is given to the ideal that is generated by $\psi_{A,y}$ in $C[x]$. In systems and control theory $K_\nu(A, y)$ is the reachable subspace of a single-input-single-output (SISO) system, and in case of a minimal realization, $\nu(A, y)$ is equal to the McMillan degree of the system; see, e.g., [2, 17].

²Our reason for preferring the terminology “Krylov space method” over “Krylov subspace method” is that the German “Krylovunterrauummethode” is even much clumsier than “Krylovraum-methode”.

---

2
Proof. The proof of Lemma 1 shows that $K_\nu(A, y)$ is an invariant subspace. Moreover, by
definition of $K_\nu(A, y)$, any invariant subspace of $A$ that contains $y$ must contain $K_\nu(A, y)$. Likewise, any polynomial $\psi$ satisfying $\psi(A)y = o$ must have $\psi_{A, y}$ as a divisor. The Jordan canonical form of the restriction of $A$ to this invariant subspace is made up of some, but in general not all of the Jordan blocks of $A$. So its minimal polynomial, which is $\psi_{A, y}$, is a divisor of $\hat{\chi}_A$.

It is important to note that for Krylov spaces the minimal polynomial $\hat{\chi}_A$ and not the characteristic polynomial $\chi_A$ matters. In particular, by Krylov space methods we cannot determine the geometric multiplicity of an eigenvalue. This follows from Lemma 3 and was the main reason for introducing block Krylov space methods for solving eigenvalue problems with multiple eigenvalues; see, e.g., [1, 3]. Note in particular that if $A$ is diagonalizable, a basis for the smallest $A$-invariant subspace is given by a minimal set of eigenvectors such that $y$ is contained in their span. There can be no two eigenvectors associated with the same eigenvalue in this set.

Lemma 4. The grade $\nu(A, y)$ satisfies

$$\nu(A, y) = \min \{ n \mid A^{-1}y \in K_n(A, y) \}.$$ 

Proof. Multiplying (1.2) by $A^{-1}$ we see that

$$A^{-1}y = -(A^{\nu-1}y + A^{\nu-2}y \gamma_{\nu-1} + \cdots + A y \gamma_2 + y \gamma_1) \frac{1}{\gamma_0},$$

so $A^{-1}y \in K_\nu(A, y)$. We cannot replace $\nu$ by some $n < \nu$ here, because this would lead to

a contradiction to the minimality of $\nu$ in (1.2), and hence, to the definition of $\nu$.

From Lemma 4 it now follows quickly that once we have constructed a basis of $K_\nu(A, r_0)$ we can find the exact solution of the linear system there. This is proved, for example, in the technical report associated with [19], but must have been known long before.

Theorem 5. Let $x_*$ be the solution of $Ax = b$ and let $x_0$ be any initial approximation of it and $r_0 := b - Ax_0$ the corresponding residual. Moreover, let $\nu := \nu(A, r_0)$. Then

$$x_* \in x_0 + K_\nu(A, r_0),$$

and $\nu$ is the smallest index for which this holds.

Proof. By Lemma 4,

$$x_* - x_0 = A^{-1}(b - Ax_0) = -A^{-1}r_0 \in K_\nu(A, r_0),$$

and $\nu$ is the smallest integer with this property.

Theorem 6. Any Krylov space solver that guarantees that residuals are linearly independent (unless zero), does not break down, and applies only one matrix-vector product with $A$ per iteration will find (in exact arithmetic) the exact solution of $Ax = b$ in exactly $\nu := \nu(A, r_0)$ iterations: $x_\nu = x_*$. 

Proof. We assume by definition that $x_\nu \in x_0 + K_\nu$, so $r_\nu \in r_0 + A K_\nu \in K_{\nu+1}$. By the minimality of $\nu$ in Theorem 5 we need at least $n := \nu$ steps to be able to find $x_*$. But because the residuals are linearly independent by assumption, $r_\nu \in K_{\nu+1} = K_\nu$ must be the zero vector since $\dim K_\nu = \nu$ only. So $\nu$ iterations are enough.
Examples for methods where Theorem 6 applies are the conjugate gradient method, the biconjugate gradient method (if it does not break down), and GMRes. These and many others have this so-called finite-termination property. In contrast, Chebyshev iteration does not have this property: the first \( \nu \) residuals it generates are linearly independent, but the \( (\nu + 1) \)th is nonzero, except under very special circumstances.

Admittedly, Theorems 5 and 6 are theoretical results that are of limited practical value. Normally \( \nu \) is so large that we do not want to spend the \( \nu \) matrix-vector products that are needed to construct a basis of \( K_\nu(A, r_0) \). We want to find very good approximate solutions with much fewer matrix-vector products. Typically, Krylov space solvers provide that; but there are always exceptions of particularly hard problems. On the other hand, there exist situations where \( \nu \) is very small compared to \( N \), and then Krylov space solvers do particularly well, because \( \nu \) iterations are enough; see, e.g., [12].

However, if we replaced the target of finding the exact solution \( x \), by the one of finding a good approximate solution, we could deduce from a correspondingly adapted notion of grade results that are truly relevant in practice. An interesting approach in this direction is due to Ilić and Turner [9], but alternatives exists too.

Historically, the conjugate gradient (CG) method of Hestenes and Stiefel [6] and the biconjugate gradient (BiCG) method of Lanczos [11] were the first methods with the finite termination property. However, it is a myth that due to this property the inventors of the CG method considered it as a direct method (like Gauss elimination) and missed its power as iterative method for sparse problems.\(^3\) Certainly, among its early users there was some initial disappointment that due to roundoff this property did not hold in practice. The problems of that time (with, say, \( N \lesssim 100 \)) typically needed up to 50\% extra iterations (without preconditioning). But this was not the reason for the CG method to have limited acceptance for nearly two decades. On reasonably well-conditioned test examples of that time the SOR method was just faster than CG, and Gauss elimination was even faster provided the matrix fitted into the memory; see Chapters III–V of [4]. The main advantage of iterative methods like CG and SOR was in any case not speed but low memory requirements. The later success of the CG method, partly initiated by Reid’s article [13], was rather due to the increasing size and complexity of the problems, and, later, the ubiquity of preconditioning.\(^4\)

2. Linear systems with multiple right-hand sides and block Krylov spaces. A nonsingular linear systems with \( s \) right-hand sides (RHSs) can be written as

\[
Ax = b \quad \text{with} \quad A \in \mathbb{C}^{N \times N}, \quad b \in \mathbb{C}^{N \times s}, \quad x \in \mathbb{C}^{N \times s}. \tag{2.1}
\]

\(^3\)For sure we know this for Stiefel and his group, which included with Hochstrasser [7] and Engeli, Ginsburg, and Rutishauser [4], some of the early users of the CG method.

\(^4\)The statement in Reid’s abstract that “... the [CG] method has several very pleasant features when regarded not as a direct method for the solution of full systems of equations but as an iterative method for the solution of large and sparse systems” may have initiated the above mentioned myth, though Reid also pointed out that “... as a direct method the algorithm is not competitive with Gaussian elimination either in respect of accuracy or number of operations. Reid’s first aim was to compare various slightly different CG algorithms with respect to computational cost, roundoff, and memory requirement. But more important is that he provided evidence that for solving “very large and sparse but reasonably well-conditioned problems” (with up to a few thousand unknowns, no preconditioning) far less than \( N \) iterations are needed, and that for these problems the method is competitive with Chebyshev iteration and SOR, but does not require an estimate of the range of the spectrum.
Most other authors choose the notation $AX = B$, but we still want to use boldface lowercase letters for the “high and skinny” $N \times s$ matrices of the unknowns and the RHSs. We will call such $N \times s$ matrices block vectors. Their $s$ columns will be distinguished by an upper index when they are addressed separately. Such columns will be written in italics, not in bold.

We gather the $s$ initial approximations for the $s$ systems in the block vector $x_0 \in \mathbb{C}^{N \times s}$ and determine the initial block residual

$$r_0 := b - Ax_0 \in \mathbb{C}^{N \times s}.$$  \hfill (2.2)

A block Krylov space solver has the property that the $n$th iterates gathered in $x_n$ are up to the shift $x_0$ linear combinations of all the columns of $r_0, Ar_0, \ldots, A^{n-1}r_0$:

$$x_n - x_0 = \sum_{k=0}^{n-1} A^k r_0 \gamma_k, \quad \text{where } \gamma_k \in \mathbb{C}^{s \times s} \quad (k = 0, \ldots, n - 1).$$  \hfill (2.3)

This leads to the following definition.\footnote{In systems and control $B_{\nu}^\square(A, y)$ is the reachable subspace of a multiple-input-multiple-output (MIMO) system \cite{17}.}

**Definition.** Given $A \in \mathbb{C}^{N \times N}$ nonsingular and $y := \left( y^{(1)} \ldots y^{(s)} \right) \in \mathbb{C}^{N \times s}$ with $y^{(i)} \neq o$ for $i = 1, \ldots, s$, the block Krylov (sub)spaces $B_n^\square (n \in \mathbb{N}^+)$ generated by $A$ from $y$ are

$$B_n^\square := B_n^\square(A, y) := \text{block span } (y, Ay, \ldots, A^{n-1}y) \subseteq \mathbb{C}^{N \times s},$$  \hfill (2.4)

where ‘block span’ is defined such that

$$B_n^\square(A, y) = \left\{ \sum_{k=0}^{n-1} A^k y \gamma_k : \gamma_k \in \mathbb{C}^{s \times s} \quad (k = 0, \ldots, n - 1) \right\}.$$  \hfill (2.5)

In this notation, (2.3) can be written as

$$x_n \in x_0 + B_n^\square(A, r_0).$$  \hfill (2.6)

Again, in practice there exist methods where $x_n$ (or some columns of it) may not exist for some $n$. On the other hand, (2.6) is too general for reflecting some essential properties of block Krylov space solvers.

From now on we will for simplicity always suppose that the assumptions of the definition hold. So, in particular, all initial residuals are assumed to be nonzero.

Each column of an element of $B_n^\square(A, y)$ is itself an element of

$$B_n := B_n(A, y) := \left\{ \sum_{i=1}^{s} \sum_{k=0}^{n-1} A^k y^{(i)} \beta_{k,i} : \beta_{k,i} \in \mathbb{C} \quad (\forall k, i) \right\} \subseteq \mathbb{C}^{N},$$  \hfill (2.7)

and this subspace of $\mathbb{C}^N$ is just the sum of the $s$ Krylov subspaces $\mathcal{K}(A, y^{(i)})$:

$$B_n(A, y) = \mathcal{K}_n(A, y^{(1)}) + \cdots + \mathcal{K}_n(A, y^{(s)}).$$  \hfill (2.8)

$B_n^\square$ is then the Cartesian product of $s$ copies of $B_n$:

$$B_n^\square = B_n \times \cdots \times B_n,$$  \hfill (2.9)

\hfill s times

\footnote{In systems and control $B_{\nu}^\square(A, y)$ is the reachable subspace of a multiple-input-multiple-output (MIMO) system \cite{17}.}
Let us return to the solution of the \( s \) linear systems \( Ax = b \). Now, \( x_0^{(i)} + B_n(A, r_0) \) is the affine space where the \( n \)th approximation \( x_n^{(i)} \) of the solution of the \( i \)th system \( Ax^{(i)} = b^{(i)} \) is constructed from:

\[
x_n^{(i)} \in x_0^{(i)} + B_n(A, r_0).
\]

(2.10)

So, we should learn more about this space.

### 3. The block grade.

Clearly, if the \( ns \) vectors \( A^k y^{(i)} \in \mathbb{C}^N \) in (2.7) are linearly independent,

\[
\dim B_n = ns.
\]

But \( \dim B_n \) can be less than \( ns \) because the sum (2.8) need not be a direct sum and because \( \dim K_n(A, y^{(i)}) < n \) may hold for some \( i \). This is where the difficulties of block Krylov space solvers come from, but also some of their merits.

Like the Krylov subspaces, the subspaces \( B_n \) and \( B_n^{\square} \) are nested:

\[
B_n \subseteq B_{n+1}, \quad B_n^{\square} \subseteq B_{n+1}^{\square}.
\]

We are going to show that, again, for sufficiently large \( n \) equality holds. Based on this fact, Schmelzer [15] introduced a generalization of the grade discussed in Section 1 to block Krylov spaces. It is defined by an adaptation of Corollary 2 and will allows us to establish a number of results that are analogous to those for the ordinary grade.

**Definition.** The positive integer \( \nu := \nu(A, y) \) defined by

\[
\nu(A, y) = \min \{ n \mid \dim B_n(A, y) = \dim B_{n+1}(A, y) \} = \min \{ n \mid B_n(A, y) = B_{n+1}(A, y) \}
\]

is called block grade of \( y \) with respect to \( A \).

Then, clearly, \( n \leq \dim B_n \leq ns \) if \( n \leq \nu(A, y) \). Moreover, in analogy to the case \( n \geq \nu \) of Lemma 1 we have then:

**Lemma 7.** For \( n \geq \nu(A, y) \),

\[
B_n(A, y) = B_{\nu(A, y)}(A, y), \quad B_n^{\square}(A, y) = B_{\nu(A, y)}^{\square}(A, y).
\]

(3.2)

**Proof.** By definition of \( \nu(A, y) \), (3.2) holds for \( n = \nu = \nu(A, y) \). Since for any of the individual Krylov spaces \( K_n(A, y^{(j)}) \) in (2.8) we have clearly \( K_{n+1}(A, y^{(j)}) = K_1(A, y^{(j)}) + AK_n(A, y^{(j)}) \) it holds likewise that \( B_{n+1}(A, y) = B_1(A, y) + AB_n(A, y) \). So, in view of the nonsingularity of \( A \) and the dimensions of the subspaces involved, \( B_1(A, y) \subseteq AB_n(A, y) = B_{\nu}(A, y) \), that is, \( B_{\nu}(A, y) \) is an invariant subspace of \( A \). So, applying \( A \) to any element of \( B_{\nu}(A, y) \) does not lead beyond this space, and this does not change if we replace \( \nu \) by \( n > \nu \). Consequently, the equality on the left side of (3.2) holds for all \( n \geq \nu \). The one on the right side follows then from (2.9).

**Remark.** Note that the inequality \( \nu(A, y^{(i)}) \leq \nu(A, y) \) need not hold for all \( i = 1, \ldots, s \). For example, let \( y^{(1)} \) and \( y^{(2)} \) be two eigenvectors of \( A \) that belong to different eigenvalues, and let \( y^{(3)} = y^{(1)} + y^{(2)} \) and \( y := \begin{pmatrix} y^{(1)} & y^{(2)} & y^{(3)} \end{pmatrix} \). Then \( \nu(A, y^{(1)}) = \nu(A, y^{(2)}) = 1 \) and \( \nu(A, y^{(3)}) = 2 \), but \( \nu(A, y) = 1 \).

On the other hand, the following inequalities hold:
Lemma 8. The block grade of the block Krylov space and the grades of the corresponding individual Krylov spaces are related by

\[ \nu(A, y) \leq \max_{i=1,\ldots,s} \nu(A, y^{(i)}) \leq \chi_A. \]  

(3.3)

Proof. The claim follows from (2.8) and (3.2). If we had \( m := \max_{i=1,\ldots,s} \nu(A, y^{(i)}) < \nu(A, y) \), it would follow that \( B_m(A, y) = B_{\nu(A, y)}(A, y) \) in contrast to the definition of \( \nu(A, y) \).

In the light of the above made remark the following result is not completely trivial:

Lemma 9. A block Krylov space and the corresponding individual Krylov spaces are related by

\[ B_{\nu(A, y)}(A, y) = K_{\nu(A, y^{(1)})}(A, y^{(1)}) + \cdots + K_{\nu(A, y^{(s)})}(A, y^{(s)}), \]  

(3.4)

and \( \nu(A, y) \) is the smallest index for which this holds.

Proof. We choose in (2.8) \( n \) larger than the indices \( \nu \) of all the spaces that appear in there and apply Lemma 7 on the left-hand side and \( s \) times Lemma 1 on the right-hand side. The minimality of \( \nu(A, y) \) is a consequence of its definition: \( B_{\nu(A, y) - 1}(A, y) \subseteq B_{\nu(A, y)}(A, y) \), so (3.4) cannot hold for \( \nu(A, y) - 1 \) if it holds for \( \nu(A, y) \).

Now we easily obtain the analog of the first part of Lemma 3:

Lemma 10. \( B_{\nu(A, y)}(A, y) \) is the smallest \( A \)-invariant subspace of \( \mathbb{C}^N \) that contains \( y^{(i)}, i = 1,\ldots,s \).

\( B_{\nu(A, y)}(A, y) \) is the smallest \( A \)-invariant subspace of \( \mathbb{C}^{N \times s} \) that contains \( y \).

Proof. We have seen in the proof of Lemma 7 that \( B_{\nu(A, y)}(A, y) \) is an \( A \)-invariant subspace. But any \( A \)-invariant subspace that contains \( y^{(i)}, i = 1,\ldots,s \), must contain \( K_{\nu(A, y^{(i)})}(A, y^{(i)}), i = 1,\ldots,s \), so by (3.4) it must contain \( B_{\nu(A, y)}(A, y) \).

Before we come to the adaptation of Lemma 4, we note the following generalization of (1.2). By definition of \( \nu = \nu(A, y) \), the columns of \( A' y \) are linear combinations of the columns of \( y, Ay, \ldots, A^{\nu-1} y \), and this does not hold for all columns of \( A^{\nu} y \) for any \( n < \nu \). That means that there are matrices \( \gamma_0, \ldots, \gamma_{\nu-1} \in \mathbb{C}^{s \times s} \), such that

\[ A' y = -y \gamma_0 - Ay \gamma_1 - \cdots - A^{\nu-1} y \gamma_{\nu-1}. \]  

(3.5)

Here, \( \gamma_0 \neq o \), because of the minimality of \( \nu \), but unfortunately we cannot be sure that \( \gamma_0 \) is nonsingular. (In fact, we did not assume that \( y \) has linearly independent columns, so \( \gamma_0 \) need not even be uniquely determined and may have rows where all elements are zero.) So, we cannot solve (3.5) easily for \( y \) and then apply \( A^{-1} \) to it. Nevertheless, by an alternative, more complicated argument we can still prove the following analog of Lemma 4.

Lemma 11. The block grade \( \nu(A, y) \) is characterized by

\[ \nu(A, y) = \min \left\{ n \mid A^{-1} y \in B^\square_n(A, y) \right\}. \]

Proof. Defining \( m := \min \left\{ n \mid A^{-1} y \in B^\square_n(A, y) \right\} \) we show first that \( m \leq \nu(A, y) \) and then that \( m \geq \nu(A, y) \). We know from Lemma 4 that \( A^{-1} y^{(i)} \in K_{\nu(A, y^{(i)})}(A, y^{(i)}) \), and
so by (3.4) we have also \( A^{-1}y^{(i)} \in \mathcal{B}_{\mathcal{V}}(A,y)(A,y) \) for all \( i \), which is the same as \( A^{-1}y \in \mathcal{B}_{\mathcal{V}}(A,y)(A,y) \). Hence, \( m \leq \nu(A,y) \).

If we knew that \( \mathcal{B}^m_{m+1}(A,y) \) is an \( A \)-invariant subspace of \( \mathbb{C}^{N\times s} \), we could conclude that \( m \geq \nu(A,y) \). By an argument similar to one in the proof of Lemma 7 we show that the former is indeed true. As \( A^{-1}y \in \mathcal{B}^m_{m+1}(A,y) \) we have \( y \in A\mathcal{B}^m_{m+1}(A,y) \subseteq \mathcal{B}^m_{m+1}(A,y) \). But \( \mathcal{B}^m_{m+1}(A,y) \) is the sum of the subspaces \( A\mathcal{B}^m_{m}(A,y) \) and \( \mathcal{B}^m_{m}(A,y) = \text{block span} \{y\} \). However, as \( y \in A\mathcal{B}^m_{m}(A,y) \) we know that \( \mathcal{B}^m_{m+1}(A,y) \subseteq A\mathcal{B}^m_{m}(A,y) \) and therefore

\[
A\mathcal{B}^m_{m}(A,y) = \mathcal{B}^m_{m+1}(A,y) .
\]

As \( A \) is an invertible linear map acting on \( \mathcal{B}^m_{m+1}(A,y) \) the dimension of \( \mathcal{B}^m_{m+1}(A,y) \) is the same as that of \( \mathcal{B}^m_{m}(A,y) \). Knowing that those subspaces form a nested sequence we end up with

\[
A\mathcal{B}^m_{m}(A,y) = \mathcal{B}^m_{m+1}(A,y) = \mathcal{B}^m_{m}(A,y) .
\]

So, by definition of \( \nu(A,y) \) or by Lemma 10 we have \( m \geq \nu(A,y) \).

Next we are looking for an analog of Theorem 5. Amazingly, we do not need Lemma 11 for its proof.

**Theorem 12.** Let \( x_* \) be the block solution of \( Ax = b \) and let \( x_0 \) be any initial block approximation of it and \( r_0 := b - Ax_0 \) the corresponding block residual. Moreover, let \( \nu := \nu(A,r_0) \). Then

\[
x_* \in x_0 + \mathcal{B}^\nu(V,A,r_0) ,
\]

and \( \nu \) is the smallest index for which this holds.

**Proof.** To prove (3.6) we just combine Theorem 5 and the relations (3.1) and (2.9). The minimality of \( \nu \) follows from Theorem 5 and Lemma 9.

Unfortunately, the generalization of the polynomial relation (1.3) and the related second part of Lemma 3 is not straightforward since we cannot write \( A' y + A^\nu y'_{\nu-1} + \cdots + A y'_{1} + y y_0 = 0 \) as \( \psi_{A,y}(A)y = 0 \). We will return to this point in Section 5.

**4. Block Krylov bases.** In the case of a single right-hand side we know from Lemma 1 that for \( n \leq \nu \) the spanning set \( r_0, Ar_0, \ldots, A^{\nu-1}r_0 \) in (1.1) is a basis of \( K_n(A,r_0) \), the so-called Krylov basis. The corresponding \( N \times n \) Krylov matrix

\[
K_n := \begin{pmatrix} r_0 & Ar_0 & \cdots & A^{n-1}r_0 \end{pmatrix}
\]

is typically very ill-conditioned, but there are situations where the basis or the matrix are used for theoretical derivations.

In the multiple right-hand side case, where we want to switch again from \( r_0 \) to an arbitrary \( y \) with \( s \) nonzero columns, the columns of the \( N \times ns \) block Krylov matrix

\[
B_n := \begin{pmatrix} y & Ay & \cdots & A^{n-1}y \end{pmatrix}
\]

are still a spanning set of \( \mathcal{B}_n(A,y) \), but they are in general no longer linearly independent. Clearly, it may happen that \( ns > N \). For example, assume \( \nu(A,y^{(1)}) = N \) and let \( y^{(2)} = A y^{(1)} \); then, for \( s = 2 \) and \( n = N \), \( B_n(A,y) \) has \( ns = 2N \) columns, but, of course, only \( N \) of them are linearly independent.

Kent [10] assumed in his treatment of block Krylov matrices that \( N \) is a multiple of \( s \) and that \( B_{N/s} \) is nonsingular. In this case we have

\[
\mathbb{C}^{N} = K_{N/s}(A,y^{(1)}) \oplus K_{N/s}(A,y^{(2)}) \oplus \cdots \oplus K_{N/s}(A,y^{(s)}) ,
\]

8
which is a very special situation. We would like to discuss the general case instead.

To get an analog of the Krylov basis in general, we may need to delete some of the columns of $B_n$, and we want to do this in such a way that for $n = 1, \ldots, \nu$, the columns of the constructed reduced block Krylov matrix $B_n^o$ are the basis vectors of a sequence of nested block Krylov bases for $B_k(A, y)$, $k = 1, \ldots, n$. This reduction corresponds to the need for deflation in block Krylov space methods. Assume $B_{n-1}^o$ has been constructed for some $n-1 < \nu$. To get $B_n^o$ we append to $B_{n-1}^o$ the columns of $A^{n-1}y$ and check for linear dependence of the new columns. For example this could be done by updating the QR decomposition of $B_n^o$, possibly using column pivoting. (In practice, one might want to use another rank-revealing QR decomposition or even the more expensive singular value decomposition.) If we detect linear dependency, we delete one or more of the columns of $A^{n-1}y$, but we do not make any changes to the columns that are inherited from $B_{n-1}^o$.

This procedure is in general not unique. If, for example, the first column of $A^{n-1}y$ appears in a linear combination that expresses the second column in terms of the first, we have the option of deleting either the first or the second column. Of course, we could opt for a specific rule, but we prefer to leave the choice open. This means that, in general, $B_n^o$ is not uniquely determined. So, there exists not just one sequence of nested block Krylov bases for $B_1, \ldots, B_n$, but a whole tree of such bases. If we denote the dimension of $B_n(A, y)$ by $d_n = d_n(A, y)$ then, clearly, any $B_n^o$ has $d_n$ columns and, for $k = 1, \ldots, n-1$, its first $d_k$ columns form a basis of $B_k(A, y)$ and are thus a possible choice for $B_k^o$.

If $d_n = N$, $B_n^o$ is nonsingular and hence its columns form a basis of $\mathbb{C}^N$. In analogy to the terminology mentioned in the Footnote 1, we suggest to call $y$ block $A$-cyclic in this case.

5. Block minimal polynomials. The minimum polynomial $\psi(A, y)$ of $y$ with respect to $A$ defined in (1.3) had the property that $\psi(A, y)y = 0$. How can we extend this to the block case? Can we also write (3.5) in terms of a polynomial?

Using the block Krylov matrix $B_o$ defined by (4.2) with $n = \nu$ we can write (3.5) as

$$A^\nu y = -B_o c, \quad \text{where} \quad c := \begin{pmatrix} \gamma_0 \\ \vdots \\ \gamma_{\nu-1} \end{pmatrix} \in \mathbb{C}^{\nu s \times s}, \quad (5.1)$$

or, if we use instead the reduced block Krylov matrix $B_o^\circ$,

$$A^\nu y = -B_o^\circ c^\circ, \quad \text{where} \quad c^\circ := \begin{pmatrix} \gamma_0^\circ \\ \vdots \\ \gamma_{\nu-1}^\circ \end{pmatrix} \in \mathbb{C}^{d_o \times s}. \quad (5.2)$$

Here, $\gamma_i^\circ \in \mathbb{C}^{(d_i+1-d_i) \times s}$, that is, these blocks of coefficients are in general of different size. While in (5.1) $B_o$ is uniquely defined, but $c$ may be non-unique, in (5.2) $B_o^\circ$ is in general nonunique, but $c^\circ$ is unique once $B_o^\circ$ has been chosen.

Given a matrix polynomial $\varphi$ of degree $m$ and with coefficients in $\mathbb{C}^{s \times s}$,

$$\varphi(t) = \beta_0 t + \beta_1 t + \cdots + t^m \beta_m, \quad (5.3)$$

it is unfortunately impossible to form $\varphi(A)y$, because the dimensions do not match. But let us recall from Kent [10] and Simoncini and Gallopoulos [16] the following
notation (which, in [16], is attributed to William B. Gragg):

\[ \varphi(A) \circ y := y \beta_0 + Ay \beta_1 + \cdots + A^m y \beta_m. \quad (5.4) \]

In this notation, and with \( \iota \) the \( s \times s \) unit matrix, (3.5) and (5.1) can be written as

\[ \psi(A) \circ y = 0, \quad (5.5) \]

where

\[ \psi(t) := \psi_{A,y}(t) := t^n \iota + t^{n-1} \gamma_{n-1} + \cdots + t \gamma_1 + \gamma_0. \quad (5.6) \]

This looks very similar to the non-block case in (1.3), but it does not reflect the possible need for deflation that lead to (5.2). The difficulty is that the coefficients \( \gamma_k \) need to be of the same size to define a matrix polynomial. We can bail out by enforcing that those rows of \( c \) that do not appear in \( c^0 \) are set to zero. Generalizing [10] we may call the corresponding (nonunique) polynomial \( \psi_{A,y} \) a matrix-valued minimal polynomial of \( y \) with respect to \( A \). It has a uniquely determined degree and a minimum number of nonzero rows in \( c \).

Matrix polynomials have been studied extensively by Gohberg, Lancaster, and Rodman [5], but the construction (5.4) seems to have received little attention.

6. Block companion matrices. Due to the special structure of \( B_\nu \), Eq. (5.1) implies that

\[ AB_\nu = B_\nu C, \quad (6.1) \]

where

\[ C := \begin{pmatrix} o & \cdots & o & -\gamma_0 \\ o & \vdots & \ddots & \ddots \\ \vdots & \ddots & \ddots & -\gamma_1 \\ o & \cdots & o & -\gamma_{n-1} \\ \iota & \cdots & \cdots & -\gamma_0 \end{pmatrix} \quad (6.2) \]

is a \( \nu s \times \nu s \) block companion matrix. Particularly beneficial is the case where \( \nu s = N \) and \( B_\nu \) is nonsingular, like under the assumption (4.3). Then \( A \) is similar to \( B_\nu \). More general results are covered in [5].

7. Conclusions. The block grade that we introduced here is not the dimension of an exhausted block Krylov space and yet it is the natural generalization of the grade of a Krylov space. As we have shown, almost all properties of the grade can be reformulated for the block grade, although alternative approaches for most of the proofs have to be used. There are further related mathematical concepts, like the block Krylov basis, the corresponding block companion matrix, and the matrix polynomial that is in a certain sense a block minimal polynomial for the starting vector that generates the block Krylov space. For all these constructions the possibility of linear dependency of columns of the block Krylov matrix — which is related to the possibility of deflation in block Krylov methods — causes extra difficulties.

There are also close connections to system and control theory, to certain areas of pure linear algebra, and to the theory of matrix polynomials, but these have only been noted in the margin here.
REFERENCES


